引文内容分析视角下图书情报领域知识计量研究刍议*

■ 毕崇武[1] 叶光辉[2] 夏立新[2]
[1] 郑州大学信息管理学院 郑州 450001 [2]华中师范大学信息管理学院 武汉 430079

摘 要：[目的/意义]目前图书情报领域的知识计量研究因无法以知识单元为计量单位，始终停留在知识载体计量层面与知识属性计量层面。提出一种从引文内容分析视角实现知识内容计量的研究思路，这是一个值得探索的研究方向。[方法/过程]在梳理不同领域的知识计量研究现状的基础上，总结图书情报领域开展知识计量研究的核心优势和关键问题，提出从引文内容分析视角开展知识计量研究的主要设想和实现路径，并展望了这一研究思路的创新应用和发展方向。[结果/结论]从引文内容分析视角开展知识计量研究能够突破图书情报领域长久以来的研究瓶颈，并真正将知识的计量对象从知识载体深入到知识内容中，实现针对知识单元的数量、质量、价值和关系计量。

关键词：知识计量 引文内容 知识单元 知识评价
分类号：G250
DOI：10.13266/j.issn.0252-3116.2021.21.017

知识计量学是刘则渊教授于1998年在“科研评价暨科学计量学与情报计量学国际研讨会”上提出的新学科设想[1]。根据定义，知识计量学是以整个人类知识体系为对象，运用对象分析和计算技术对社会的知识（生产、流通、消费、累积和增殖等）能力和知识的社会关系（组织形式、协作网络、社会建制等）进行综合研究的一门交叉学科[2]。受研究视角和时代因素的双重影响，不同学科、不同领域的学者们对“知识计量”概念的内涵和外沿具有不同认识。从研究对象来看，现有知识计量研究集中在科学学、经济学、管理学、计算机科学和图书情报学五个领域[3]。

图书情报领域的知识计量活动大致经历了以知识载体（如期刊、专利等文献）为研究对象、以知识属性（如主题词、关键词等信息）为研究对象和以知识内容（如知识本体、知识单元等知识）为研究对象的三个发展阶段。早在1996年，马费成教授便指出情报学取得突破在微观上需要解决两个关键问题：一是知识信息的表达和组织必须从物理层次的文献单元向认识层次的知识单元或情报单元转换；二是知识信息的计量必须从语法层次向语义和语用层次发展[4]。在学者们把目光从文献计量、信息计量推进到知识计量的过程中，他们始终希望能够实现从间接的知识计量上升到直接的知识计量，并由此提出了最小、独立的知识单位——知识单元（或知识元）；然而限于当时的技术、方法等因素，他们很难将知识单元与文献单元、信息单元进行有效分割。因此，尽管学者们在过去从理论层面揭示了知识单元的概念，指出了知识单元对于知识计量的重要性，但是依然无法实现真正的知识内容计量。长久以来，知识单元构建成为制约图书情报领域深入开展知识计量研究的真正瓶颈。

近些年，随着开放获取运动的兴起，获取文献资源的全文本数据变得十分便利。与此同时，以章节结构划分、引用位置获取等为代表的文本结构识别技术，和以文献主题提取、引用强度计算等为代表的文本内容挖掘技术，也为基于学术文献中文本数据的分析和计量研究提供了更加广阔的空间和可能。过去学者们对如何揭示知识之间的序化联系只能采用手工标注的方式，然而随着知识图谱技术发展，现有研究不仅能够做到自动抽取知识关联，而且能够利用可视化图谱形象地展示领域知识的体系架构、发展脉络和前沿动向。由此可见，环境的改变和技术的发展已经为图书情报领域开展知识计量研究提供了全新的契机，以往的制

* 本文系国家社会科学基金重大项目“新时代我国文献信息资源保障体系重构研究”（项目编号：19ZDA345）研究成果之一。
作者简介：毕崇武（OCRID：0000-0001-7874-285X），讲师，E-mail：767818984@qq.com；叶光辉（OCRID：0000-0001-8111-5034），副教授，硕士生导师；夏立新（OCRID：0000-0002-4162-2282），教授，博士生导师。
收稿日期：2021-04-21 修回日期：2021-06-08 本文起止页码：97-107 本文责任编辑：杜杏叶

约条件已经有所改变,因此有必要对图书情报领域的知识计量研究进行重新思考。

1 图书情报领域的知识计量研究现状与演变

知识是一个复杂的概念,对知识进行计量也是一个艰难的过程。以研究对象的演进路径为脉络,图书情报领域的知识计量研究经历了知识载体计量、知识属性计量和知识内容计量三个阶段。这一动态演变过程展现了学者们开展知识计量研究时的艰难探索和逐步深入。

1.1 知识载体计量阶段

以存储知识的载体——文献单元,作为计量单元,得益于日益完善、成熟的文献计量学指标、方法和评价体系。在这一阶段,学者们将文献视作是知识,通过对文献进行定量化研究来间接计量知识的主要特征,其间借鉴的极具代表性且具有重大影响的研究成果有 F. J. Cole 等[5]基于出版物数量的时间分布曲线,比较了某一学科在不同时期的研究重点及发展变化;P. L. K Gross 等[6]通过统计分析期刊论文的参考文献,获取某一领域内的核心期刊;E. Garfield[7]提出的利用引文索引来检索科技文献的研究思路;P. O. Seglen[8]论述的影响因子应用于期刊评价的局限性;H. Small[9]利用共被引概念测量被引文献集群对某一领域的贡献程度;S. C. Bradford[10]发现的某一领域论文在期刊中的分布规律;A. J. Lotka[11]发现的文献作者与其生产效率之间的数量关系;J. E. Hirsch[12]提出的量化文献作者科研水平的 H 指数;以及文献的增长规律、老化规律、引用规律、齐氏定律等等。

总体说来,知识载体计量阶段聚焦于知识的数量和价值特征,通过知识的物化对物化的知识进行测度与评价[13],但是存在以下问题:其一,文献与知识并不能完全等同,文献中必然包含了知识,但是这种做法无法计量文献中包含了多少知识;其二,忽略了知识质量与知识价值间的区别,仅将知识质量等同于文献价值;其三,知识关系计量方面较为薄弱。

1.2 知识属性计量阶段

以揭示知识的属性——信息单元,作为计量单元,得益于伴随信息技术发展而产生的用于知识表达及传播的新型非文献资源。在这一阶段,学者们用信息刻画知识,在对文献单元进行结构化分解的基础上,通过信息单元及其组合来标记、表达知识的内外特征;并产生了大量关于知识计量的理论性研究成果。典型研究如文庭孝等曾对知识计量的研究对象与内容、基本单元、学科渊源和主要方法展开系统综述[14],并详细地阐述了知识计量的研究维度和计量单元[15-16];余以胜等[17]曾对知识计量研究的发展历程、测度内容和主要困难展开深入探讨。事实上,这些研究都认识到了知识载体计量的局限性,也指出了知识内容计量的重要性,并提出了以知识单元为计量单位的设想,但是并没有将知识内容计量实现。

总体说来,知识属性计量阶段是一个过渡时期或中间时期。它既没有克服知识载体计量的固有缺陷,也没有实现知识内容计量的最终目标,且存在以下问题:其一,信息单元只是控制和处理知识载体的手段,并不具备完全独立的现实意义,通常不能脱离知识载体而单独存在,仍然需要利用知识载体定位知识内容;其二,该阶段研究只是探究了知识线索,虽然这在一定意义上可以视为是考虑了知识关系,但是却无法精确化描述知识链路。

1.3 知识内容计量阶段

以描述知识的内容——知识单元,作为计量单元,得益于文本挖掘技术的广泛应用,尤其是得益于针对文献资源的知识识别技术和知识抽取技术。在这一阶段,有三个方面的研究对实现知识内容计量具有重要意义。一是知识单元理论的深入发展,促使学术界对知识单元有了一个更为清晰的认知,使知识单元不再是一个虚拟、模糊的概念;例如不少研究设计了知识元的描述规则及其服务模式,并在一定程度上实现了知识元的抽取与关联[18-19]。二是文献结构功能的识别与划分研究实现了从文献资源中定位知识内容,例如章节划分[20]、段落划分[21]和功能划分[22]等。三是文献内容特征的识别与提取研究实现了从文献资源中分离知识内容,例如词汇识别[23]、句法识别[24]和语义识别[25]等。

在此基础上,随着数据处理能力的日益提升,加之公开文献资源的日益丰富,越来越多的研究工作开始使用学术文献中的文本数据进行计量分析,其中被广泛使用的文本数据是引文内容。引文内容蕴含了文献作者从施引文献的文本数据中抽取的被引文献的知识内容,通过引文内容分析能够揭示施引文献和被引文献之间的创新关系以及被引原因。因此,从引文内容分析视角开展知识计量研究既能够深入到知识载体中的知识单元,又能够保留知识单元间的复杂关系。目前引文内容分析研究可以划分为 4 个方面:一是知识

单元与实体抽取，即从学术文献中抽取知识的最小表示单元及其关联实体，常见的抽取方法有基于规则、基于统计和混合策略[24, 26－27]；二是引用位置与程度分析，通过识别引文内容在施引文献中的出现位置及分布情况，对简单统计引用频次的传统引文分析进行扩展和延伸[28－29]；三是引用行为与动机分析，依据引文内容的描述方式猜测施引者发生引用行为的真实目的，其中部分研究融入了社会学与心理学研究方法[30－31]；四是，引用情感与类型识别，利用引文内容的语义特征获取施引者对被引文献的主观情感，进而将其拓展至被引文献的学术评价中[32－33]。

总体说来，现有研究已经为实现知识内容计量提供了可能性，未来的发展趋势也必然是进一步实现知识内容的准确计量。然而在当下，仅有个别学者意识到现有研究已经为基于知识单元的新型知识计量范式的产生与发展创造了条件[34]，部分研究仅仅注意到基于全文数据的文献计量学分析悄然兴起[35－36]，另有一些研究即使大谈“知识计量”也并未做到真正的知识内容计量。事实上，无论是知识载体计量阶段、知识属性计量阶段还是知识内容计量阶段，学者们的初衷均是希望对知识本身展开测度，只是限于认知水平和技术手段，在无法精准地对知识本身进行直接计量时，只能粗略地对知识本身进行间接计量。在这个不断探索的过程之中，也衍生出了一系列极为重要的计量指标和计量方法。尽管以笔者的眼光来看，这些指标和方法并没能实现对知识本身的直接计量，但是它们的研究角度、设计理念、使用方式已经为知识计量研究的深入发展奠定了极为重要的研究基础。

2　图书情报领域的知识计量研究独具特色

知识计量是一个综合性、跨领域的研究难题，目前尚未取得突破性进展的根本原因在于知识本身的无形性。尽管如此，图书情报领域的学者们经过长期摸索，已经在知识计量的理论与问题研究、方法与工具研究、应用与实证研究等方面产出了独具一格的研究成果。

2.1　其他领域的知识计量研究

知识经济时代的提出与知识经济学的兴起，让学者们认识到有必要将科学计量学、文献计量学、信息计量学和经济计量学等研究方法与范式结合起来，从宏观和微观上对知识的生产和应用、投入和产出、存量和流量、分配和转移、价值和价格等开展广泛的跨学科综合研究[17]。因此，除图书情报领域的学者们之外，其他领域的学者们也从各自的需求出发，针对知识本身及其价值的计量问题，形成了相对独立的理论与方法体系[15]。

在科学学领域，学者们以科技成果的计量、分析、评价和管理为目的，探究科技成果中的知识及其载体的数量特征、发展规律和结构关系，以此揭示出科学技术的发展规律和科学研究的活动规律。在研究过程中，学者们以科技成果为主要研究对象，进而拓展到与科技成果相关的科研人员、科研机构、地区或国家。

在管理学领域，学者们把知识作为一种投入资源考察其产出、效率和效益，并在知识资产评估方面取得了较大进展。特别是在知识分类的问题上，管理学学者们将知识划分为可编码的显性知识和不可编码的隐性知识。其中，隐性知识的测度和计量问题，以及隐性知识向显性知识的转化问题，成为管理学界开展知识计量研究的关键问题。

在经济学领域，学者们从知识的经济价值视角来测度知识对社会经济发展的影响，其核心在于应用知识经济学的理论和方法，对知识经济和知识产业中的知识贡献率展开测度。他们把知识作为一种独立的生产要素纳入企业或社会的生产过程，通过考察和分析这种新的经济增长模式，度量经济体中的知识投入、知识存量、知识流量、知识产出等等。

在计算机科学领域，学者们针对显性知识的表达、理解、处理和存储等问题展开了广泛且深入的研究，在知识挖掘、知识发现、知识图谱、知识仓库、知识系统等方面取得了重大突破，特别是在一定程度上实现了文本、图像、音视频等资源的知识抽取、知识关联、知识服务，为这些资源的知识计量研究提供了可靠的技术手段和方法工具。

2.2　图书情报领域开展知识计量研究的核心优势

虽然不同领域的学者们从不同视角直接或间接地开展了知识计量研究，但是由于研究目的和研究意义各不相同，现有研究较为零碎、分散且不系统。除此之外，一些领域开展知识计量研究存在着难以攻克的瓶颈。例如，在科学学领域，尽管科技成果由法定机关认可，避免了研究对象的所属范围问题，但是基础研究成果、应用研究成果和发展工作成果等不同性质的科技成果转化与评估问题，极大地束缚了以科技成果为研究对象开展知识计量的研究工作。在管理学和经济学领域，无论是从知识管理视角还是知识经济视角，开展知识计量研究的最大难点在于如何衡量隐性知识的流

转问题。在计算机科学领域,现有研究侧重于知识的获取与应用,但是缺少相对成熟的知识计量研究范式。

与之相比,图书情报领域开展知识计量研究具有得天独厚的优势,主要表现在三个方面:

(1)研究对象明确且为显性知识。尽管图书情报领域的知识计量研究经历了三个不同的发展阶段,但是学者们始终围绕着文献这一知识载体开展知识计量研究。有所差异的仅仅是知识计量单元的逐渐演变。事实上,无论是以文献单元、信息单元还是知识单元作为基本计量单位,它们均是人类通过一定的方法和手段、运用一定的意义表达和记录体系,记录在载体上的有价值的显性知识。这恰恰回避了难以直接量化处理的隐性知识。

(2)知识元研究已应用到实践。早在 20 世纪 70 年代,图书情报领域的专家就指出知识的控制单位需要从文献单元深入到知识单元,期望能够直接对文献的逻辑内容进行知识组织和分析,以便找到知识创造过程中的联接节点,从而深入揭示知识的有机结构[37]。数十年间,该领域学者们从理论方法、技术模型以及应用平台等角度展开了广泛的探索与研究,并在知识元理论与述评、知识元表述与建模、知识元抽取与实现等方面取得了显著成绩。近些年,随着认知理论的发展和自然语言处理能力的提高,图书情报领域基本上实现了自动抽取文献资源的全文本知识单元,也真正做到了针对知识内容的组织与服务。

(3)拥有相对成熟的研究范式。在图书情报领域,采用计量学方法对知识载体、知识内容、知识活动及其影响等进行定量的研究工作由来已久。经过数十年的沉淀,学者们逐渐将研究视角从基于知识载体的知识计量转移到基于知识内容的知识计量上。与此同时,图书情报领域中的文献计量、信息计量和网络计量等与知识计量具有密切亲缘关系的计量学科已经发展到了一个成熟阶段,其技术手段、工具和方法均可以为图书情报领域开展知识计量研究奠定坚实基础。知识计量研究是以往这些计量研究的推广、深化和发展。

2.3 图书情报领域开展知识计量研究的关键问题

早期图书情报领域开展知识计量研究最核心的问题在于如何从文献中获取最小、独立的知识计量单元。尽管学者们很早就有了以知识单元作为知识计量单元的设想,但是限于对知识单元概念定义的不清晰,特别是知识单元构建的困难,图书情报领域的知识计量研究在很长一段时间里并没有得到深入发展,现有研究只是围绕着知识载体和知识属性特征开展,没有做到对知识内容的直接计量。在某种程度上,过去图书情报领域的知识计量研究只是借鉴了文献计量学、信息计量学、科学计量学等研究方法和研究范式,除了研究思路与研究视角十分超前之外,在研究内容上与已有计量学的研究内容无异。当前,图书情报领域开展知识计量研究的核心问题仍然是知识单元的构建问题,其先决条件是明确文献中的知识单元是什么,以及如何抽取知识单元。除此之外,还应当考虑如下关键问题:

(1)知识单元的数量计量问题。众所周知,文献是有历史意义或研究价值果的一种资源,其中必然蕴含着一定数量的知识单元。那么根据知识单元的概念,一篇文献中的知识单元数量如何计算?文献中的知识单元如何分布?文献之间的知识单元数量有何差异?

(2)知识单元的质量计量问题。知识单元可视作是最小、独立的知识产品。作为一种产品,知识单元就有满足规定需要和潜在需要的特征和特性的总和。那么如何衡量知识单元的实际质量呢?同一篇文献中的不同知识单元是否具有相同的知识质量?知识质量与知识价值之间的关系又是什么?在以往的研究中,部分学者将知识的质量与知识的影响混为一谈,利用知识载体在引文网络中的被引次数衡量知识的质量。事实上,被引次数衡量了知识载体作为知识产品的使用次数,是一种使用价值的表现形式。

(3)知识单元的价值计量问题。根据政治经济学中使用价值和价值的定义,我们很难衡量一个知识单元的价值,也无法准确给出一个知识单元的使用价值,但是我们可以利用知识单元的使用次数评判不同知识单元之间使用价值的大小。这也由此引发出一些值得思考的研究问题。如何获取知识单元的使用次数?知识单元是否只能被完整的使用?

(4)知识单元的关系计量问题。知识单元间存在着十分复杂的逻辑关联,因此在知识计量过程中也应当区分知识单元间的知识连接类型。在相同文献中,知识单元之间的连接关系会影响知识单元的数量计量结果;在不同文献中,知识单元间的连接关系会影响知识单元的价值计量结果。此外,还需要特别注意的是,相同文献中的知识单元之间与不同文献中的知识单元之间,其知识单元之间的知识连接计算方式和种类范围也应有所差异。

3　从引文内容分析视角开展知识计量研究的主要设想

除明确的研究对象、适宜的技术手段和成熟的研究范式之外，引文网络为图书情报领域的学者们提供了一个支持大数据分析的研究平台。在以往的图书情报领域研究中，利用引文网络开展知识计量研究屡见不鲜，其根本原因在于引文网络契合了图书情报领域开展知识计量研究的研究对象与研究范式。然而，在过去基于引文网络的知识计量研究中，学者们限于知识单元的构建困境，并没有深入到知识内容开展知识计量研究，只是宏观、表面地测量了知识载体或知识属性。鉴于此，本文从引文内容分析视角出发，利用针对文献资源的全文本挖掘技术，提出一个图书情报领域开展知识计量研究的新的设想，如图 1 所示：

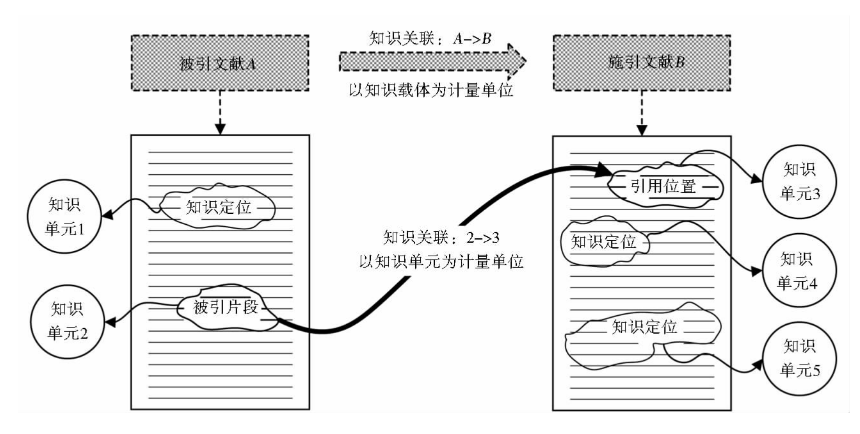


图 1　以知识单元为计量单位的知识计量研究实现路径

文献中蕴含了表达观点、定义、理论、方法等以半结构化或非结构化形式存在的知识单元，需要首先利用知识的描述规则定位文献中知识单元的具体位置，然后依据事先设计的描述框架对相应位置的文本内容进行知识抽取与知识存储，最后基于文献间的引用关系，将施引文献在引用位置上的知识单元与被引文献在被引片段上的知识单元关联起来，构建以知识单元及其之间连接关系为基础的知识网络。

以知识网络为基础，可以从引文内容分析视角开展知识计量研究。①文献中的知识单元数量决定了知识节点中的知识存量；②被引片段上的知识单元决定了被引文献提供的知识内容；③引用位置上的知识单元决定了施引文献接收的知识内容；④通过比较被引片段和引用位置上的知识单元间的差异，能够刻画出知识节点间的引用强度，而引用强度决定了知识节点间的知识流量；⑤通过综合分析引用位置、引用情感和引用动机等信息，能够定义知识节点间的引用类型，而引用类型决定了知识单元间隐含的特定联系。鉴于上述思考，本文设计了针对 2.3 小节中关键问题的解决思路。

3.1　文献中的知识单元数量计量思路

从文献中抽取知识单元是计算知识数量的基础。目前知识抽取有基于统计和基于规则两条技术路线。文献中的知识多由复杂度较高的句子组成，采用数理统计与机器学习方法通常难以适应，因此采用规则识别与模式匹配方法较为适宜。鉴于现有的认知理论发展水平和自然语言处理能力，本文设计了如图 2 所示的知识单元数量计量步骤。

首先，需要解决知识单元的刻画与描述问题。为了从训练文本中获得一个高质量的知识单元描述规则，以此为知识单元的自动化抽取奠定基础，可以采取机器与人工相结合的半自动化方式。在此过程中，需要尽可能保证训练语料覆盖范围广泛且具有代表性、文本语句划分完整、句子结构及语法成分准确、领域知识主题词表完善。

然后，需要解决知识单元的识别与抽取问题。通过对目标文献开展语句划分、文本分词和主题词过滤等自然语言处理工作，识别并提取出可能含有知识单元的文本片段；依据生成的知识单元描述规则，将文本片段的依存句法序列与描述规则的句法依存序列相互匹配，发现隐藏在半结构化或非结构化文本片段中知识单元。

最后，需要从多个维度测度与分析知识单元的数量特征，以便全面揭示不同知识主体的知识单元总量、流量、增长、分布、结构等等。

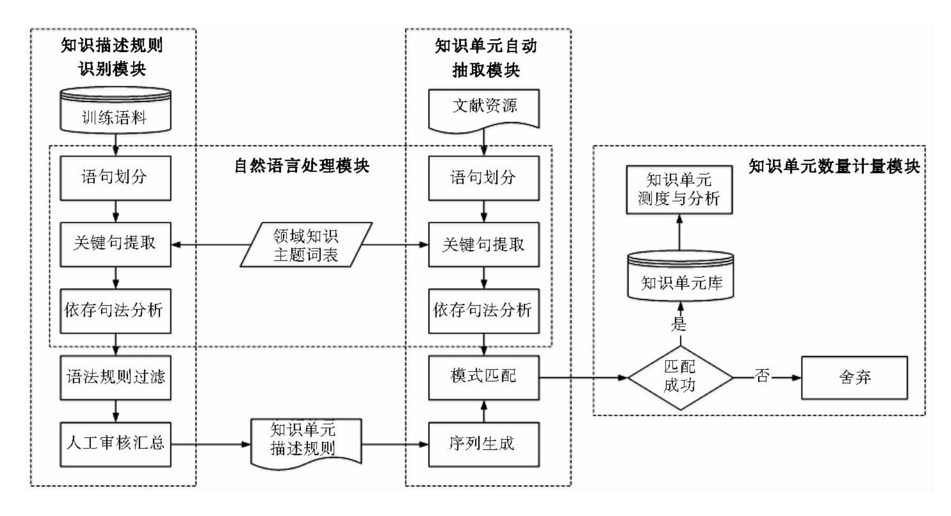


图 2　知识单元的数量计量实现路径

3.2　文献中的知识单元质量计量思路

遵循质量之本质在于满足和超越预期这一经典论述，知识的质量实际上就是知识评价者感知知识特性符合其自身预期的程度[37]。知识不像普通产品，不能够以直接的定量参数表示其质量，但是其质量可以利用质量的特性或特征来间接反映。因此，在很大程度上，对知识的质量评价应该是主观的。这种主观性源于不同评价者实施评价时依据的评价维度之间存在差异。从文献单元到信息单元再到知识单元，随着评价对象的产品属性越来越抽象，质量的内涵就越来越复杂，所需的评价维度也越来越丰富。鉴于此，本文在理论层面选用多维度评价标准来全面揭示与知识质量这一复杂概念相适应的图景，详见图 3，其优势在于多维度分解能够使知识质量的概念由抽象变得具体，与此同时也能够降低误判风险。

在实践层面，尽管不同的评价模型在评价维度上可能存在一定差别，但是整体上应该有规律可循。可以从信息维度上输入诠释知识质量特征的信号，例如作者、机构、期刊、项目等，从知识属性方面衡量知识单元的质量；从过程维度上输入揭示知识生产过程的文本，评价生产知识单元的各个环节是否合理、环节之间的联系是否正确；从本体维度上输入展现知识内在构面与情景构面的体系框架，评价知识单元与既有知识体系之间的逻辑关系；从应用维度上输入实证过程和复用条件，评价知识单元解决实际问题的应用效果。在此基础上，可以在每个评价维度下设立更为具体的评价指标，进而借助同行评议法针对每项评价指标开展专家评分工作，以此实现针对知识单元质量特征的量化计算和归一化处理。

此外，还可以借助机器学习中的相关算法实现对知识单元质量的快速评估。利用文本挖掘技术可以直接从输入数据中提取各评价指标对应的质量特征 x。例如，假设知识单元的质量特征 x 和预测结果 p 之间符合某种回归关系，那么首先通过模型训练可以优化评价模型中各评价指标的实际参数，然后选取优化算法不断缩小预测结果 p 与专家评分 y 之间的误差数值，最终可以实现同行评议结果与机器学习结果达到一致。

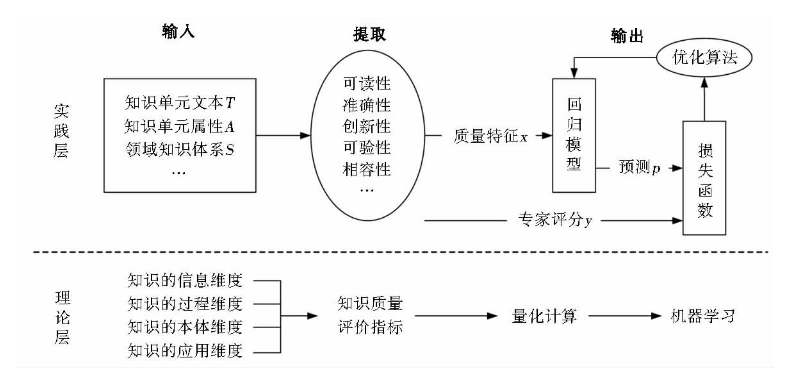


图 3　知识单元的质量计量实现路径

3.3 文献中的知识单元价值计量思路

知识的价值在于流动,其价值不仅仅取决于知识自身的质量,更取决于知识的流动深度和广度。文献中的知识单元之间并不是孤立存在的。通常一篇文献要建立在多篇文献的基础上,通过知识单元间的引用行为来实现。这一客观事实既符合了科学本身的发展规律,又体现了研究活动中知识单元间的继承性和变化性。一般说来,被引文献是施引文献的研究基础,被引片段对施引文献具有指导作用。因此,可以视为被引片段中蕴含的知识单元流向了施引文献中对应的引用位置。鉴于此,本文设计了依据知识单元流动状况评价知识单元价值大小的研究思路,如图 4 所示。

在知识流动测量模块,以施引文献和被引文献构成的引用关系为基础,以知识单元为计量单位,测量单一三元组中的知识单元流量。首先,需要从施引文献中识别引用位置,进而抽取引文内容;其次,比对引文内容,从被引文献中识别并抽取被引片段;然后,依据知识单元描述规则,生成施引、被引知识单元;最后,计算知识单元间的相似程度,并以此为依据衡量单次引用过程中的知识流量。在计算相似度时,需要考虑知识单元间的细分主题、句法结构和文本内容,通过特征融合设计权重综合计算最终数值。

在知识价值测量模块,以引文数据库提供的引用关系为数据基础,以知识单元在引文网络中的流动状况为评价依据,评估知识单元的价值大小。首先,从引文数据库中抽取知识单元间的流动关系,即知识单元由被引文献流向了施引文献;其次,构建以知识单元为网络节点的异质、加权网络;然后,借助知识单元之间的知识流量,并融入揭示其关系类型的外部数据,全面分析知识单元在知识网络中的位置与角色;最后,融合内容相似性算法和角色重要性算法,对互相依存、互相影响的知识单元进行价值排序。

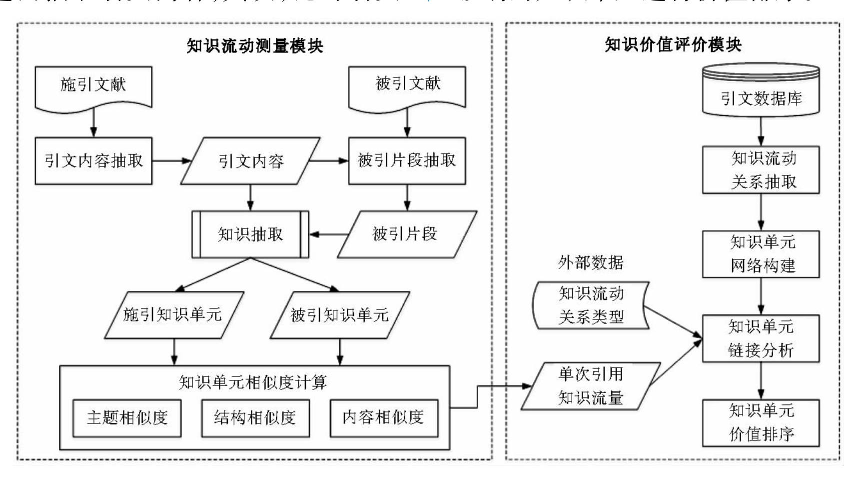


图 4 知识单元的价值计量实现路径

3.4 知识单元间的关系类型计量思路

知识单元之间存在着各种关联,揭示和利用这些关联是知识组织、知识管理和知识服务的起点。知识单元之间的关系类型,实际上是隐藏在知识单元之间,可以被理解和序化的关联种类。依据引用关系或知识属性直接关联的知识单元之间,其关系类型是显而易见的;但是不直接关联的知识单元间的关系类型,也可以通过中介知识单元进行挖掘。从知识单元的来源来看,知识单元间的关系类型需要考虑相同文献中的知识关联和不同文献中的知识关联 2 个方面。由此,本文设计了如图 5 所示的知识单元间关系计量研究思路。

在相同文献中,需要首先在知识单元分类的基础之上,将知识单元划分为不同的类型,并依据知识描述框架对其进行结构化存储;然后综合考虑专家意见和训练结果,设定合理的知识关系类别及划分标准,以此为基础指导知识单元之间关系特征的提取;最后以关联向量存储知识单元间不同的关联特征,并分别计算不同关联向量中的语义相似程度,最终融合所有关联向量的计算结果,综合评判知识单元之间的关系类型。

在不同文献中,需要重点依据引文内容分析结果划分知识单元间的关系类型。引用关系搭建起了知识单元间最直接的知识关联,但是这种关联蕴含的关系类型十分复杂。尽管现有技术可以将施引文献中的引文内容与被引文献中的被引片段对应起来,但是在实践过程中很难将引用行为的初始动机与最终结果统一起来。因此,需要全面分析引文内容与被引片段中对应的知识单元,综合考虑引用位置、引用情感、引用目

的、引用强度等计算结果，从多种角度、多种视野、多种层面来认识和发现不同文献中知识单元间的关系类型。

总体说来，上述两个方面的侧重点有所不同。在相同文献中的知识单元之间，由于知识的生产者相同，应主要考虑知识单元间的交叉关系、对立关系、并列关系等等；在不同文献中的知识单元之间，由于存在知识的“遗传”和“变异”，应主要考虑知识单元间的等同关系、种属关系、对立关系等等。此外，上述研究思路仅仅涉及由两个知识单元构成的二元关系，然而知识单元间的关联关系还具有一定的传递性。若从三方关系结构角度考虑，则知识网络可拆分为可分析的微观单元并揭示出更为丰富结构特征[39]。

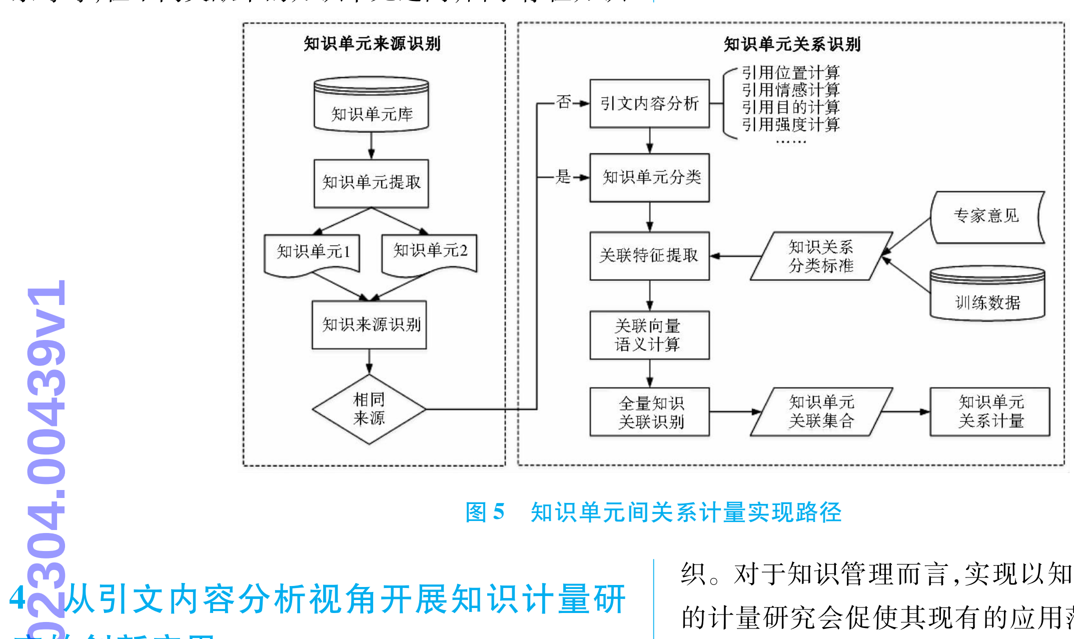


图 5　知识单元间关系计量实现路径

4　从引文内容分析视角开展知识计量研究的创新应用

知识计量是一切知识活动的重要基础，在知识的生产、获取、组织和呈现过程中具有广泛的应用。通过上述思考可以看出，借助当前的认知理论发展水平和自然语言处理能力，并基于“五计学”现有的研究方法和理论，基本能够实现从引文内容分析视角开展知识计量研究。这一崭新的研究视角也将带来更为引人注目的应用场景，足以促进未来知识计量研究的深化和拓展。纵观国内外研究趋势，未来知识计量研究的主要应用集中在三个方面：

4.1　在知识管理中的应用方向

过去不少情报学家曾提出图书情报领域的研究对象是知识的论断，但是在很长一段时间里并没有受到学术界的足够重视[40]。回顾图书情报领域的研究历程可以发现，图书情报领域的工作流程始终沿着“事实——数据——信息——知识——情报/智能”这条信息链进行逐级提炼升华，且当前研究集中在将信息加工为知识这一环节[41]。本文的研究思路能够将知识管理的控制单位从文献单元和信息单元过渡到知识单元，这也由此可以从间接的知识管理上升到直接的知识管理，并可以对知识内容进行有效地识别、处理与组织。对于知识管理而言，实现以知识单元为基本单位的计量研究会促使其现有的应用范畴发生重大调整：①可以服务于知识数量管理，能够测量出任何知识所有者的知识总量、知识增量，从而弥补现有研究仅能利用指数增长模型及其变型测量文献的总量与增长量，或利用信息熵、摩尔定律测量信息的总量与增长量，但是涉及知识总量与增长量测量时仅有难以定量计算的布鲁克斯知识方程式可以借鉴的缺陷；②可以服务于知识质量管理，能够评估知识的学术质量、开展知识的质量检测与质量控制，不再以期刊论文等知识载体为对象开展科研成果的管理与评价活动；③可以服务于知识价值管理，能够以最小、独立的知识单位为基础评估知识的投入与产出情况，进而计算出知识及其相关产品的使用价值、社会影响以及科技进步贡献率；④可以服务于知识关系管理，能够将单一的引用关系拓展至复杂的知识关联，进而揭示出知识间隐藏的、可理解的、可使用的内在逻辑。

4.2　在科学评价中的应用方向

图书情报领域开展科学评价源于 20 世纪中叶兴起的文献计量和引文分析。通过借助文献资源的各种特征数量，采用数学与统计的分析方法，可以描述、评价和预测科学技术的发展现状和发展趋势，并对参与科学活动过程中的相关实体开展量化评估。在漫长的

研究历程中,不少学者也曾指出现有的理论、方法和指标,虽然对当今科学评价起到了不可替代的积极作用,但是仍然存在一些不足,不能完全照搬;应当根据具体状况进行必要分析、合理使用、有效借鉴、甚至改造,才能使分析评价结果更加客观、合理和科学[42]。

对于科学评价而言,实现以知识单元为基本单位的计量研究可以被应用于与科学活动有关的各个领域。学术文献中包含了十分丰富的知识单元,以知识单元作为评价对象,将引用位置、强度、功能、情感、动机等特征相结合,能够开展真正脱离依赖学术期刊的学术评价活动,从根源上遏止学术评价中"唯论文""SCI 至上"等"以刊评文"的不良导向,具体包括 4 个方面:①可将知识单元计量结果融入到科研成果评价,包括对著作、期刊、论文和专利等的地位和作用评价;②可将知识单元计量结果融入到科研人员评价,包括对学术影响力、科研绩效率等的定性与定量评价;③可将知识单元计量结果融入到科研领域评价,评价一个学科或研究领域的发展阶段、现状、水平、前景,及其知识结构与知识内容间的相关度等;④可将知识单元计量结果融入到科研机构评价,大到一个城市、区域甚至国家,小到一个部门或学校,均可测度其科研产出的数量、质量、效果和应用。

4.3 在知识服务中的应用方向

经过多年的理论探讨,图书情报领域当前仍然难以开展高效、精准的知识服务,"知识饥渴"与"信息泛滥"的对峙局面依然存在。从知识资源角度来看,知识虽然多分布于文献资源之中,但是其目前的组织方式仍然多以粗粒度的文献单元为基本单元,无法匹配用户多粒度、层次化的知识需求。与此同时,描述相同研究对象或科研问题的知识也通常分布在不同的学科之中,相互之间存在一定的学科隔阂,无法揭示不同知识之间的关联[43]。这些现象均阻碍了学界和业界为终极用户提供更精良的知识服务体验。

对于知识服务而言,实现以知识单元为基本单位的计量研究能够厘清知识储备及其体系结构,针对不同用户提出的实际需要,选择与其约束条件相符的计量结果,实现更深层次的专业化知识服务:①能够实现细粒度的知识服务,促使以知识载体(如图书、期刊、专利等)、信息(如主题、数据、关键词等)为单元的知识服务方式向以知识内容为单元的知识服务方式转变。例如,通过对文献资源当中的知识单元进行挖掘提取和整理剖析,向用户提供某一实例、数值、概念、定律等人类知识的最小单位。②能够实现组合式的知识服务,促使基于文献、按学科的知识服务方式向基于知识内容的跨学科知识服务方式转变。例如,利用知识关系的计量结果构筑跨学科、跨领域的客观知识体系,向用户提供融合多层次知识的概念认知、解决措施、操作流程等知识组配方案。

5 总结与展望

自 20 世纪 90 年代起,为了迎接知识经济时代,知识被看作是一种社会竞争中的战略资源和经济资源,受到了社会各界前所未有的关注和重视。近些年,尽管科学计量学、文献计量学、信息计量学等研究,均在其各自领域内涉及到了对于知识载体或知识属性的计量,但是由于研究目的和研究意义不同,这些研究未能深入到知识内容开展位于知识科学高度的系统研究,也难以令人信服地揭示和阐明知识在现代社会发展中的重要作用。本文从引文内容分析视角,设想出一条图书情报领域开展知识内容计量的可行路径,并从文献中的知识单元数量计量、质量计量、价值计量和关系计量 4 个方面,探讨了开展知识内容计量的基础问题及其解决措施,最后畅想了从引文内容分析视角实现知识内容计量后的应用场景。

从目前情况来看,知识计量仍然是一个介于多个学科之间的研究方向,具有跨学科的研究特征。这要求图书情报领域开展知识计量研究时,不仅要借鉴其他学科的理论、方法与技术,还要将这些融会贯通到本学科内,使之上升到一个新的高度,深化出一个更为完善、符合内在逻辑的研究体系与研究范式。因此,缩小学科之间差距,清除阻隔,综合和整合各学科领域正在进行的不同层面、不同视角的知识计量研究尤为重要。例如,科学学领域的科研绩效评价与管理方法,管理学领域的显性与隐性知识管理理论,经济学领域的宏观与微观经济价值视角,计算机科学领域的知识挖掘与存储技术,以及教育学领域和信息科学领域关于知识概念的认识,均会对图书情报领域开展知识计量学研究具有较大的参考价值。

但更为重要的是,图书情报领域开展知识计量研究还必需内生出一些新的、其他学科不曾具备或不太完善的理论与方法。在漫漫的历史进程中,新理论、新方法和新技术对图书情报领域的影响是十分深远的;但是在任何时候都应该根据自身实际,结合自身特色,做到有所为和有所不为,不应一味求全贪大,否则就会舍本逐末,失去自身特色。总而言之,在科学研究交叉与融合的时代,在机遇与挑战并存的环境之中,图书情

报领域的学者们应该做到在传承与创新、坚守与拓展中找准定位,开展有图书情报特色的知识计量研究。

参考文献:

[1] 刘则渊. 赵红州与中国科学计量学[J]. 科学学研究, 1999, 17(4): 104-109.
[2] 刘则渊, 刘凤朝. 关于知识计量学研究的方法论思考[J]. 科学学与科学技术管理, 2002, 22(8): 5-8.
[3] 邱均平, 文庭孝, 宋艳辉, 等. 知识计量学[M]. 北京: 科学出版社, 2014.
[4] 马费成. 情报学的进展与深化[J]. 情报学报, 1996, 15(5): 338-344.
[5] COLE F J, HULME E W. The history of comparative anatomy - a statistical analysis of the literature [J]. Science progress, 1917 (11): 578-596.
[6] GROSS P L K, GROSS E M. College libraries and chemical education[J]. Science, 1927, 66 (1713): 385-389.
[7] GARFIELD E. Citation indexes for science: a new dimension in documentation through association of ideas [J]. Science, 1955, 122(3159): 108-111.
[8] SEGLEN P O. Why the impact factor of journals should not be used for evaluating research [J]. British medical journal, 1997, 314 (7079): 497-502.
[9] SMALL H. Co-citation in the scientific literature: A new measure of the relationship between two documents [J]. Journal of the American Society for Information Science and Technology, 1973, 24 (4): 265-269.
[10] BRADFORD S C. Sources of information on specific subjects[J]. Engineering, 1934, 137: 85-86.
[11] LOTKA A J. The frequency distribution of scientific productivity [J]. Journal of the Washington academy of sciences, 1926, 16 (12): 317-323.
[12] HIRSCH J E. An index to quantify an individual's scientific research output[J]. The National Academy of Sciences of the USA, 2005, 102(46): 16569-16572.
[13] 文庭孝. 知识计量与知识评价研究[J]. 情报学报, 2007, 26 (5): 670-676.
[14] 文庭孝, 刘晓英, 梁秀娟, 等. 知识计量研究综述[J]. 图书情报知识, 2010, 26(1): 95-101.
[15] 文庭孝, 刘璇. 论知识计量研究的维度[J]. 图书情报知识, 2013, 29(3): 103-109.
[16] 文庭孝. 知识计量单元的比较与评价研究[J]. 情报理论与实践, 2007, 30(6): 731-736.
[17] 余以胜, 张洋. 知识的计量与评价研究[J]. 图书情报工作, 2008, 52(11): 18-21.
[18] 王忠义, 黄容, 郑鑫, 等. 数字图书馆多粒度集成知识服务研究[J]. 情报学报, 2019, 38(2): 143-158.
[19] 毕崇武, 王忠义, 宋红文. 基于知识元的数字图书馆多粒度集成知识服务研究[J]. 图书情报工作, 2017, 61(4): 115-122.
[20] 黄永, 陆伟, 程齐凯. 学术文本的结构功能识别——基于章节内容的识别[J]. 情报学报, 2016, 35(3): 293-300.
[21] 黄永, 陆伟, 程齐凯, 等. 学术文本的结构功能识别——基于段落的识别[J]. 情报学报, 2016, 35(5): 530-538.
[22] 陆伟, 黄永, 程齐凯. 学术文本的结构功能识别——功能框架及基于章节标题的识别[J]. 情报学报, 2014, 33(9): 979-985.
[23] 陆伟, 李鹏程, 张国标, 等. 学术文本词汇功能识别——基于BERT 向量化表示的关键词自动分类研究[J]. 情报学报, 2020, 39(12): 1320-1329.
[24] 化柏林. 学术论文中方法知识元的类型与描述规则研究[J]. 中国图书馆学报, 2016, 42(1): 30-40.
[25] 温浩. 科技文摘创新点语义识别与分类方法研究[J]. 情报学报, 2019, 38(3): 249-256.
[26] YING D, MIN S, HAN J, et al. Entitymetrics: measuring the impact of entities[J]. Plos one, 2013, 8(8): e71416.
[27] 索传军, 盖双双. 知识元的内涵、结构与描述模型研究[J]. 中国图书馆学报, 2018, 44(4): 54-72.
[28] HU Z G, CHEN C M, LIU Z Y. Where are citations located in the body of scientific articles? a study of the distributions of citation locations[J]. Journal of informetrics, 2013, 7(4): 887-896.
[29] HU Z G, LIN G G, SUN T A, et al. Understanding multiply mentioned references[J]. Journal of informetrics, 2017, 11(4): 948-958.
[30] WILLETT P. Readers' perceptions of authors' citation behaviour [J]. Journal of documentation, 2013, 69(1): 145-156.
[31] BORNMANN L, DANIEL H. What do citation counts measure? a review of studies on citing behavior[J]. Journal of documentation, 2008, 64(1): 45-80.
[32] CATALINI C, LACETERA N, OETTL A. The incidence and role of negative citations in science[J]. Proceedings of the National Academy of Sciences, 2015, 112(45): 13823-13826.
[33] SULA C A, MILLER M. Citations, contexts, and humanistic discourse: toward automatic extraction and classification[J]. Literary and linguistic computing, 2014, 29(3): 452-464.
[34] 孙震, 冷伏海. 基于知识元的新型科学计量范式探析[J]. 情报学报, 2017, 36(6): 555-564.
[35] 卢超, 章成志, 王玉琢, 等. 语义特征分析的深化——学术文献的全文计量分析研究综述[J]. 中国图书馆学报, 2021, 47 (2): 110-131.
[36] 胡志刚, 章成志. 悄然兴起的全文计量分析[J]. 图书馆论坛, 2021, 41(3): 1-11.
[37] 温有奎. 知识元挖掘[M]. 西安: 西安电子科技大学出版社, 2004.
[38] 周波, 张芳. 知识市场情境的知识质量评价模型[J]. 管理学报, 2013, 10(10): 1535-1544.
[39] 董克, 刘德洪, 江洪. 基于三方关系组的引用网络结构分析[J]. 情报理论与实践, 2010, 33(11): 50-53.

[40] 邱均平，段宇锋. 论知识管理与图书情报学的变革[J]. 中国图书馆学报，2003，29(2)：4－8.
[41] 马费成，李志元. 新文科背景下我国图书情报学科的发展前景[J]. 中国图书馆学报，2020，46(6)：4－15.
[42] 苏新宁. 文献计量学与科学评价中有关问题思考[J]. 图书与情报，2013，32(1)：79－83.
[43] 王忠义，夏立新，李玉海. 基于知识内容的数字图书馆跨学科多粒度知识表示模型构建[J]. 中国图书馆学报，2019，45(6)：50－64.

作者贡献说明：
毕崇武：论文拟题，论文框架设计、撰写与修改；
叶光辉：参与论文框架设计，提出思路；
夏立新：提出修改意见，论文审定。

Rethinking Knowledge Measurement in Library and Information Science from the Perspective of Citation Content Analysis

Bi Chongwu[1]　Ye Guanghui[2]　Xia Lixin[2]
[1] School of Information Management, Zhengzhou University, Zhengzhou 450001
[2] School of Information Management, Central China Normal University, Wuhan 430079

Abstract: [**Purpose/significance**] The research of knowledge measurement in Library and Information Science can not take the knowledge element as measurement unit, and always stays at measurement level of knowledge carrier and knowledge attribute. This paper puts forward a research idea of knowledge content measurement from the perspective of citation content analysis, which is a research direction worthy of exploration. [**Method/process**] This paper summarized the core advantages and key issues of knowledge measurement in Library and Information Science on the basis of combing research status of knowledge measurement in different fields, put forward the main ideas and implementation path of knowledge measurement from the perspective of citation content analysis, and looked forward to the innovative application and development direction of this research idea. [**Result/conclusion**] From the perspective of citation content analysis, the research of knowledge measurement can break through the long-standing limitation in Library and Information Science, truly penetrate the measurement unit from the knowledge carrier to the knowledge content, and realize the quantity, quality, value and relationship measurement of knowledge element.
Keywords: knowledge measurement　citation context　knowledge element　knowledge evaluation